UNIVERSITI TUNKU ABDUL RAHMAN
Department of Mathematics and Actuarial Science

CONTENTS

4 Empirical Bayes Parameter Estimation

4.1 Introduction

In the previous Chapters, the parameters (μ, v, a) needed to determine the credibility weighted premium are all given. In this chapter, we will study two methods to estimate these parameters.

Empirical Credibility

- Nonparametric estimation: This is based on unbiased estimators of μ, v and a,
- Semiparametric estimation: $f(x|\theta)$ is parametric (usually Poisson or Geometric) but the $\pi(\theta)$ is nonparametric.

4.2 Nonparametric Estimation

4.2.1 Bühlmann Straub Model:

For each policyholder $i, 1 \le i \le r$ (and $r > 1$), we have observations $X_i = (X_{i1}, X_{i2}, \ldots, X_{i,n_i})$ of loss per exposure unit corresponding to exposures $m_i = (m_{i1}, m_{i2}, \ldots, m_{i,n_i})$ and $n_i > 1$. This means that $m_{ij}X_{ij}$ is the aggregate loss for period (or unit) j from policyholder i. Let $m_i = \sum_{j=1}^{n_i} m_{ij}$ be the total exposure for policyholder i. Be careful to distinguish between X_{ij} (a rate) and $m_{ij}X_{ij}$ (an amount).

Estimation of BühlmannStraub parameters μ, v, and a:

- **STEP 1.**
  Calculate the sample mean $\bar{x}_i$ and biased σ_i^2 (σ^2 in TI-30).

$$\bar{x}_i = \frac{1}{m_i}\sum_{j=1}^{n_i} m_{ij}x_{ij};$$

$$\sigma_i^2 = \frac{\sum_{j=1}^{n_i} m_{ij}(x_{ij} - \bar{x}_i)^2}{m_i}$$

then the (unbiased) sample variance for each policyholder.

$$\hat{v}_i = \frac{1}{n_i - 1}\sum_{j=1}^{n_i} m_{ij}(x_{ij} - \bar{x}_i)^2 = \frac{m_i}{n_i - 1}\sigma_i^2$$

- **STEP 2.**
  Calculate the weighted average of the sample variances $\hat{v}_i$ with weights $n_i - 1$. If the policyholders all have the same number of periods of exposure, this is exactly the average of the sample variances.

$$\hat{v} = \frac{\sum_{i=1}^{r}(n_i - 1)\hat{v}_i}{\sum_{i=1}^{r}(n_i - 1)}$$

- **STEP 3.**
  Calculate

$$\hat{\mu} = \bar{x} = \frac{\sum_{i=1}^{r} m_i \bar{x}_i}{m}$$

and biased

$$\sigma^2 = \frac{\sum_{i=1}^{r} m_i(\bar{x}_i - \bar{x})^2}{m}$$

then,

$$\hat{a} = \frac{\sum_{i=1}^{r} m_i(x_i - \bar{x})^2 - (r-1)\hat{v}}{m - m^{-1}\sum_{i=1}^{r} m_i^2} = \frac{m\sigma^2 - (r-1)\hat{v}}{m - m^{-1}\sum_{i=1}^{r} m_i^2}$$

- The estimator $\hat{a}$ may be negative. If it is negative but small in absolute value, the authors suggest setting it to zero and $Z_i = 0$ for all i. The estimator for all risks is $\hat{\mu} = \bar{x}$ in this case.
- After estimating the Bühlmann parameters, we estimate a given client's credibility premium based on its own experience as

$$\hat{Z}_i \bar{x}_i + (1 - \hat{Z}_i)\hat{\mu}$$

where

$$\hat{k} = \frac{\hat{v}}{\hat{a}}$$

and

$$\hat{Z}_i = \frac{m_i}{m_i + \hat{k}} = \frac{m_i \hat{a}}{m_i \hat{a} + \hat{v}}$$

Example 1. Past data on a portfolio of group policyholders are given below.
Estimate the BühlmannStraub credibility premiums to be charged to each group member in year 4.

Policyholder		Year 1	Year 2	Year 3	Year 4
Claims	1	$---$	20,000	25,000	$---$
No. in group		$-$	100	120	110
Claims	2	19,000	18,000	17,000	$---$
No. in group		90	75	70	60
Claims	3	26,000	30,000	35,000	$---$
No. in group		150	170	180	200

203.79, 225.82, 183.1

Example 2 (T4Q1).

EMBY-Q18 Past data on two group policyholders are available and are given in the following table. Determine the estimated total credibility premium to be charged to the first group in year 4.

	Policyholder	Year 1	Year 2	Year 3	Year 4
Total Claims	1	-	10950	12150	-
No. in Group		-	90	140	140
Total Claims	2	21400	25700	22650	-
No. in Group		60	150	170	250

Example 3 (T4Q2).

EMBY-Q39 An insurance company has for five years insured three different types of risk. The number of policies in the j^{th} year for the i^{th} type of risk is denoted by m_{ij} for $i = 1, 2, 3$ and $j = 1, 2, 3, 4, 5$. The average claim size per policy over all five years for the i^{th} type of risk is denoted by $\bar{X}_i$. The values of m_{ij} and $\bar{X}_i$ are tabulated below.

	Number of policies					Mean claim size
Risk type i	Year 1	Year 2	Year 3	Year 4	Year 5	$\bar{X}_i$
1	19	23	25	29	38	885.0
2	49	61	69	56	43	709.0
3	45	31	70	93	112	906.0

The insurance company will be insuring 30 policies of type 1 next year and has calculated the aggregate expected claims to be 26073.66 using the assumptions of Empirical Bayes method. Calculate the expected annual claims next year for risks 2 assuming the number of policies will be 45.

Example 4 (T4Q3).

EMBY-Q30 For four group policyholders, the number of exposures in each group for years 1 and 2 are:

	Year 1	Year 2
I	50	58
II	39	49
III	–	25
IV	–	5

Empirical Bayes non-parametric methods are used to assign credibility. You are given:

$$\sum_{i,j} m_{ij}(x_{ij} - \bar{x}_i)^2 = 15,000$$

$$\sum_{i} m_i(\bar{x}_i - \bar{x})^2 = 53,000$$

Determine the credibility assigned to group 1.

Example 5 (T4Q4).

EMBY-Q34 Group 1 and 2 have each been observed for 4 years. Group 3 has been observed for 3 years. Each of these groups had m_{ij} members in year j and average losses of x_{ij} per member, where i is the group number. Summary statistics are

$$\sum_j m_{ij} = \begin{cases} 300 & i = 1 \\ 800 & i = 2 \\ 200 & i = 3 \end{cases}$$

$$\sum_j m_{ij}x_{ij} = \begin{cases} 99,999 & i = 1 \\ 310,000 & i = 2 \\ 310,000 & i = 3 \end{cases}$$

$$\sum_j m_{ij}x_{ij}^2 = \begin{cases} 37,499,333 & i = 1 \\ 123,133,333 & i = 2 \\ 620,666,667 & i = 3 \end{cases}$$

Using empirical Bayes nonparametric estimation, determine the credibility factor for Group 1.

4.2.2 Credibility-Weighted-Average

The method which preserves total losses
Because exposures may vary by group, credibility factors Z will also vary by group. This means that the mean of the predictive estimates will not be the mean of the distribution. To avoid this problem, instead of using $\hat{\mu} = \bar{x}$, we can define $\hat{\mu}$ as the credibility weighted average:

$$\hat{\mu} = \frac{\sum_{i=1}^{r} \hat{Z}_i \bar{x}_i}{\sum_{i=1}^{r} \hat{Z}_i}$$

Example 6 (T4Q5).

EMBY-Q20 You are given the following experience for two insured groups:

Group		Year			
		1	2	3	Total
1	Number of members	9	10	6	25
	Average loss per member	91	85	96	89.8
2	Number of members	30	30	25	85
	Average loss per member	120	116	126	120.35
Total	Number of members				110
	Average loss per member				113.41

Determine the nonparametric Empirical Bayes credibility premium for group 1, using the method that preserves total losses.

4.2.3 Bühlmann model-Estimating μ, v, and a

$n_i = n$ and $m_{ij} = 1$ for all i and j.

STEP 1.
Calculate the sample mean $\bar{x}_i$.
Then, the unbiased estimator of v_i is simply

$$\hat{v}_i = \frac{\sum_{j=1}^{n} (x_{ij} - \bar{x}_i)^2}{n-1} = s_i^2$$

for each risk.

STEP 2.

The unbiased estimator of v is

$$\hat{v} = \frac{1}{r} \sum_{i=1}^{r} \hat{v}_i$$

STEP 3.

The unbiased estimators of μ and a are

$$\hat{\mu} = \bar{x}$$

and

$$\hat{a} = s^2 - \frac{\hat{v}}{n}$$

where

$s^2 = \frac{\sum_{i=1}^{r} (\bar{x}_i - \bar{x})^2}{r-1}$

is the unbiased estimate of the variance of the sample means.

Example 7 (T4Q6).

IBY-Q06b Three individual policyholders have the following claim amounts over four years:

Policyholder	Year 1	Year 2	Year 3	Year 4
X	130	150	120	150
Y	240	200	240	100
Z	370	375	280	400

Using the nonparametric empirical Bayes procedure, estimate the pure premium for the coming year for Policyholder Y.

Example 8. You are given the following experience for losses for 2 policyholders over a period of 3 years:

Policyholder	Year 1	Year 2	Year 3
A	10	12	14
B	$10 - x$	$12 - x$	$14 - x$

Empirical Bayes nonparametric methods are used to assign credibility to this experience.

Determine the range of values of x for which the credibility factor assigned by empirical Bayes nonparametric methods is greater than zero.

Example 9. The table below gives the annual aggregate claim amounts for the past 5 years for 3 risks.

	Year				
	1	2	3	4	5
Risk 1	233	247	271	215	209
2	184	219	177	227	203
3	191	169	176	168	201

(a) Using the normal/normal model, with parameter values $\mu = 216$, $v = 400$, $a = 200$, estimate the pure premium for the coming year for each of the three risks.

(b) Using nonparametric empirical Bayes method, estimate the pure premium for the coming year for each of the three risks.

(c) Comment on the differences between the answers to parts (a) and (b).

Example 10 (T4Q7).

EMBY-Q19 An insurer has data on losses for four policyholders for seven years. X_{ij} is the loss from the ith policyholder for year j. You are given:

$$\sum_{i=1}^{3}\sum_{j=1}^{7}(X_{ij}-\bar{X}_i)^2 = 46.7;$$
$$\sum_{i=1}^{3}(\bar{X}_i-\bar{X})^2 = 4.0$$

Calculate the Buhlmann credibility factor for an individual policyholder using nonparametric empirical Bayes estimation.

Example 11 (T4Q8).

EMBY-Q37 The table below shows the annual aggregate claim statistics for three individual policyholders over four years. The annual aggregate claim for risk i, in year j, is denoted by x_{ij}.

Policyholder	$\bar{x}_i = \frac{1}{4}\sum_{j=1}^{4} x_{ij}$	$s_i^2 = \frac{1}{3}\sum_{j=1}^{4}(x_{ij}-\bar{x}_i)^2$
X	1,185	24,700
Y	2,060	410,400
Z	3,734	110,056

Using the nonparametric empirical Bayes procedure, Calculate the value of the credibility factor.

4.2.4 Data with only One Policyholder:

In this special case of $r = 1$, one policyholder, this method can still be applied if the unconditional mean (or manual premium) μ is given. The formulas in this case are simply

$$\bar{x} = \frac{\sum_{j=1}^{n} m_{1j}x_j}{m_1}$$

$$\hat{v} = \frac{\sum_{j=1}^{n} m_{1j}(x_{1j}-\bar{x})^2}{n-1} = \frac{m\sigma_1^2}{n-1}$$

$$\tilde{a} = (\bar{x}-\mu)^2 - \frac{\hat{v}}{m_1}$$

Example 12. Past data on one policyholder are available and are given in the following table. Determine the estimated credibility premium to be charged in year 3 if the manual rate is 500 per year.

	Year 1	Year 2
Claims	200	400

350

Example 13 (T4Q9)**.**

EMBY-Q17 For a group policyholder, we have the following data available:

	Year 1	Year 2	Year 3
Total Claims	4000	7000	-
No. in Group	10	15	20

If the manual rate per person is 450 per year, estimate the total credibility premium for year 3 using the nonparametric method.

Example 14 (T4Q10)**.**

EMBY-Q09 The following data are available for a group policyholder:

	Year 1	Year 2	Year 3
Total claims	28,790	37,030	–
Number in groups	80	120	140

The manual rate per exposure is 490 per year. Estimate the total credibility premium for year 3 using empirical Bayes non-parametric methods.

4.3 Empirical Bayes Semi-Parametric

If $n_i = 1$, $i = 1, \cdots, r$, the nonparametric estimation method discussed in the previous section does not work ($v_i = 1$). In this section, we can apply two special semiparametric estimation methods:

- $f(x|\theta)$ has a Poisson distribution.
- $f(x|\theta)$ has a Geometric distribution.

4.3.1 Poisson Model

In this case, we have

$$E(X) = E[E(X|\Theta)] = E[\Theta] = \mu$$

$$v = E[V(X|\Theta)] = E[\Theta] = \mu$$

$$\hat{v} = \hat{\mu} = \bar{x}$$

$$V(X) = E[V(X|\Theta)] + V[E(X|\Theta)]$$
$$= v + a$$
$$= \mu + a$$

$$\hat{a} = \hat{V}(X) - \hat{\mu} = s^2 - \bar{x}$$

$$\hat{k} = \frac{\hat{v}}{\hat{a}}$$

where

$$\bar{x} = \frac{\sum_{i=1}^{r} x_i}{r}$$

$$s^2 = \frac{\sum_{i=1}^{r}(x_i - \bar{x})^2}{r-1}$$

Example 15 (T4Q11).

EMBY-Q10 For a group of auto policyholders, you are given:

- The number of claims for each policyholder has a conditional Poisson distribution.
- During Year 1, the following data are observed:

Number of Claims	number of Policyhoders
0	13010
1	7070
2	2070
3	490
4	160
5+	0

A randomly selected policyholder had 3 claims in Year 1. Determine the semiparametric empirical Bayes estimate of the number of claims in Year 2 for the same policyholder.

Example 16 (T4Q12).

EMBY-Q11 The following information comes from a study of robberies of convenience stores over the course of a year:

- X_i is the number of robberies of the i^{th} store, with $i = 1, 2, \ldots, 410$.
- $\sum X_i = 40$
- $\sum X_i^2 = 210$
- The number of robberies of a given store during the year is assumed to be Poisson distributed with an unknown mean that varies by store.

Determine the semiparametric empirical Bayes estimate of the expected number of robberies next year of a store that reported 1 robberies during the studied year.

Example 17 (T4Q13).

EMBY-Q01 You are given:

- During a 6-year period, 23,320 policies had the following claims experience:

Total Claims in Year 1 - Year 6	Number of Policies
0	13,400
1	5,820
2	2,280
3	1,700
4	120

- The number of claims per year follows a Poisson distribution.
- Each policyholder was insured for the entire 6-year period.

A randomly selected policyholder had 2 claims over the 6-year period. Using semiparametric empirical Bayes estimation, determine the Buhlmann estimate for the number of claims in Year 7 for the same policyholder.

Example 18 (T4Q14).

EMBY-Q12 For a large sample of insureds, the observed relative frequency of claims during an observation period is as follows:

Number of Claims	Relative Frequency of Claims
0	62.0
1	27.0
2	8.0
3	1.0
4	2.0
5+	0

Assume that for a randomly chosen insured, the underlying conditional distribution of number of claims per period given the parameter Θ is Poisson with parameter Θ. Given and individual who had c claims in the observation period. The semi empirical Bayesian estimate of the expected number of claims that the individual will have in the next period is 0.6504. Determine c.

Example 19 (T4Q15).

EMBY-Q31 The number of claims submited by seven policyholders over three months is shown in the following table:

	January	February	March
A	2	0	1
B	1	1	1
C	2	0	1
D	2	2	3
E	3	3	2
F	2	1	2
G	1	1	2

The number of claims for the following year is estimated using empirical Bayes semiparametric methods. It is assumed that each policyholder's annual claims follow a Poisson distribution. Unbiased estimators are used for the expected value of the process variance and the variance of hypothetical means.

Calculate the credibility projection of the annual number of claims for policyholder A.

Example 20 (T4Q16).

EMBY-Q23 You are given the followings:

- The number of losses arising from $m + 52$ individual insureds over a single period of observation is distributed as follows:

Number of Losses	Number of Insureds
0	m
1	32
2	20
3 or more	0

- The number of losses for each insured follows a Poisson distribution, but the mean of each such distribution may be different for individual insureds.
- The variance of the hypothetical means is to be estimated from the data.

Determine all values of m for which the estimate of the variance of the hypothetical means will be greater than 0.

4.3.2 Geometric Model

In this case we assume $(X|\Theta)$ follow a Geometric distribution.

$$E(X) = E[E(X|\Theta)] = E[\Theta] = \mu$$

$$v = E[V(X|\Theta)] = E(\Theta(1+\Theta)] = \mu + E(\Theta^2)$$

$$\Longrightarrow E(\Theta^2) = v - \mu$$

$$V(X) = v + a = v + v - \mu - \mu^2$$

$$s^2 = 2\hat{v} - \bar{x} - \bar{x}^2$$

$$\Longrightarrow \hat{v} = \frac{s^2+\bar{x}+\bar{x}^2}{2}$$

$$\hat{a} = \hat{v} - \bar{x} - \bar{x}^2$$

$$\Longrightarrow \hat{a} = \frac{s^2-\bar{x}-\bar{x}^2}{2}$$

$$\hat{k} = \frac{\hat{v}}{\hat{a}} = \frac{s^2+\bar{x}+\bar{x}^2}{s^2-\bar{x}-\bar{x}^2}$$

Example 21 (T4Q17).

EMBY-Q36 For a group of auto policyholders, you are given:

- The number of claims for each policyholder has a conditional Geometric distribution.
- During Year 1, the following data are observed:

Number of Claims	number of Policyhoders
0	14990
1	2580
2	1190
3	185
4	90
5+	0

A randomly selected policyholder had 0 claims in Year 1. Determine the semiparametric empirical Bayes estimate of the number of claims in Year 2 for the same policyholder.

Example 22 (T4Q18).

MBY-Q35b You are given:

- During a 6-year period, 2,041 policies had the following information and summary statistics:
  - X_i is the number of claims of the i^{th} policyholders, with $i = 1, 2, \ldots, 2,041$.
  - $\sum_{i=1}^{2,041} X_i = 767$
  - $\sum_{i=1}^{2,041} X_i^2 = 1,459$
- The number of claims per year follows a Geometric distribution.
- Each policyholder was insured for the entire 6-year period.

A randomly selected policyholder had 3 claims over the 6-year period. Using semiparametric empirical Bayes estimation, determine the Buhlmann estimate for the number of claims in Year 7 for the same policyholder.